Proteome Profiling of L3 and L4 *Anisakis simplex*

Development Stages by TMT-based Quantitative

Proteomics

Robert Stryński^{1*}, Jesús Mateos², Santiago Pascual², Ángel F. González², José M. Gallardo², Elżbieta Łopieńska-Biernat¹, Isabel Medina², Mónica Carrera^{2^}

¹University of Warmia and Mazury, Faculty of Biology and Biotechnology,
Department of Biochemistry, Olsztyn, Poland

²Spanish National Research Council (CSIC), Marine Research Institute (IIM),
Department of Food Technology, Vigo, Pontevedra, Spain

RUNNING TITLE HEAD: TMT-based proteomics of L3 and L4 of *Anisakis simplex*

***CORRESPONDING AUTHOR:** Robert Stryński, M. Sc.

University of Warmia and Mazury, Faculty of Biology and Biotechnology, Department
of Biochemistry, Olsztyn, Poland

Phone: +48 895234720

Fax: +48 895352015

E-mail: robert.stryinski@uwm.edu.pl

^CO-CORRESPONDING AUTHOR: Dr. Mónica Carrera

Spanish National Research Council (CSIC), Marine Research Institute (IIM),
Department of Food Technology, Eduardo Cabello 6, 36208, Vigo, Pontevedra, Spain.

Phone: +34 986231930

Fax: +34 986292762

E-mail: mcarrera@iim.csic.es

ABSTRACT

Anisakis simplex is a parasitic nematode that can cause anisakiosis and/or allergic reactions in humans. The presence of invasive third-stage larvae (L3) in many different consumed fish species and the fourth-stage larvae (L4) in marine mammals, where L3 can accidentally affect to humans and develop as far as stage L4. World Health Organization and food safety authorities aim to control and prevent this emerging health problem. In the present work, using Tandem Mass Tag (TMT)-based quantitative proteomics we analyzed for the first time the global proteome of two *A. simplex* development stages, L3 and L4. The strategy was divided into four steps: (a) protein extraction of L3 and L4 development stages, (b) high intensity focused ultrasound (HIFU)-assisted trypsin digestion, (c) TMT-isobaric mass tag labeling following by high-pH reversed-phase fractionation, and (d) LC-MS/MS analysis in a LTQ-Orbitrap Elite mass spectrometer. A total of 2443 different proteins of *A. simplex* were identified. Analysis of the modulated proteins provided the specific proteomic signature of L3 (i.e. pseudocoelomic globin, endochitinase 1, paramyosin) and L4 (i.e. neprilysin-2, glutamate dehydrogenase, aminopeptidase N). To our knowledge, this is the most comprehensive dataset of proteins of *A. simplex* for two development stages (L3 and L4) identified to date.

Keywords: *Anisakis simplex*, anisakiosis, proteome, quantitative proteomics, tandem mass tag (TMT), L3/L4 development stage

SIGNIFICANCE

A. simplex is a fish-borne parasite responsible for the human anisakiosis and allergic reactions around the world. The work describes for the first-time the comparison of the proteome of two *A. simplex* stages (L3 and L4). The strategy is based on four steps: (i) protein extraction, (ii) ultra-fast trypsin digestion under High-Intensity Focused Ultrasound (HIFU), (iii) TMT-isobaric mass tag labeling followed by high-pH reversed-phase fractionation and (iv) peptide analysis using a LTQ-Orbitrap Elite mass spectrometer. The workflow allows to select the most modulated proteins as proteomic signature of those specific development stages (L3 and L4) of *A. simplex*. Obtained stage-specific proteins, could be used as targets to control/eliminate this parasite and in future eradicate the anisakiosis disease.

1. INTRODUCTION

Anisakis simplex (*A. simplex*) is a fish-borne parasitic nematode that can causes anisakiosis and/or allergic reactions in humans ¹. It is characterized by a complex life cycle involving several hosts, where adult stages live in marine mammals. Its paratenic hosts are sea fish or cephalopods ^{1, 2}. *Anisakis*, particularly third- ~~and fourth~~-stage larvae (L3 ~~and L4~~) from cold-blooded fish ~~and warm-blooded marine mammals~~ respectively, can accidentally affects to humans and develop to L4 stage^{3, 4}. The entire development cycle of *A. simplex* depends on the temperature of the water⁵. Due to continuous climate change and rising temperatures, *A. simplex* extends its range, resulting appearance of this species in the seas and oceans where it has not previously been found before. That might have a direct effect on the growing number of many other fish species infections ^{6, 7, 8}. The risk of anisakiosis is mainly originate by the consumption of raw or undercooked marine fish or cephalopods parasitized with L3 larvae. Occasionally, L3 moults into L4 larvae in humans, but they do not progress to the adult stage ^{9, 10}. These larvae can elicit a parasitic infection of the digestive tract and can occasionally affect other organs, causing erosive and hemorrhagic lesions, ascites, and perforations leading to granulomas. To date, the only effective treatment for anisakiosis is the endoscopic removal of live larvae and the symptomatic treatment. Additionally, in sensitive humans, *A. simplex* larvae can cause allergic reactions ^{11, 12}. The symptomatology of this allergy arises within 60 minutes of ingestion and comprise rash, abdominal pain, vomiting, diarrhea and respiratory distress¹³. In the most severe cases, anaphylaxis shock can be potentially fatal ^{14, 15}. In the cases of allergenic patients, a fish free diet is the unique recommended treatment.

The number of reported cases of anisakiosis continues to increase. A total of 20,000 new cases of the disease were reported in 2010 worldwide¹⁶. According to the quantitative risk assessment model, the prevalence of anisakiosis in Europe will increase to 7500-8500 cases per year¹⁷. Anisakiosis is also an economic problem for fisheries due to the negative impact on consumer confidence resulting decrease in demand on the market for potentially parasitized fish species^{10, 18, 19}. Based on Scientific opinion of the European Food Safety Authority (EFSA)¹⁶, there are not Anisakis free zone, thus the European Union (EU) promulgated rules (Regulations 853/2004, 1276/2011) to control this risk²⁰. In this sense, to avoid problems of anisakiosis recommends that fish products should have to be frozen at -20°C for 24 h or at -35°C for not <15h or heating at temperatures >60°C for 1 min to kill any live nematodes in the foodstuffs^{16, 20}. As a food control method, a new targeted proteomics strategy published by our group that is based on the monitoring of several anisakids peptide biomarkers using mass spectrometry (MS) achieves the fast detection of anisakids in any food product in less than two hours²¹.

In the last years, Proteomics methods have gained more acceptance among the food scientific community^{22, 23, 24}. Carrera et al.^{21, 25} has provided new applications in the food control area using targeted proteomics methodologies for the rapid detection of Anisakidae in fishery products. However, still there is a lack of data and information regarding to those proteins, which have a crucial role in the developmental stages (L3 and L4) of this parasitic nematode.

According to that, we chose high-resolution tandem mass tag (TMT)-based quantitative proteomics to analyze global proteomic dynamics in *A. simplex*, between two developmental stages L3 and L4. Multiplex quantification using TMT 10-plex increasing the confidence in both the identification and quantification of the proteins.

The numbers of identified peptides and proteins in shotgun proteomics experiments using TMT 10-plex is highly satisfactory and the precision on the level of peptide–spectrum matches and protein level dynamic range are more acceptable than in other isobaric mass tags ^{25, 26, 27}. Each mass-tagging reagent has the same nominal mass (isobaric) and chemical structure composed of an amine-reactive NHS-ester group, a spacer arm and a mass reporter. For each sample, a unique reporter mass in the low mass region of the MS/MS spectrum is used to measure relative protein expression levels during peptide fragmentation. Each of 10 TMT labels has different structure and isotopes position. TMT labels have different numbers and combinations of ¹³C and ¹⁵N isotopes in the mass reporter through what each peptide in tested sample have monoisotopic mass differences on a scale of 126 to 131 Da (126, 127C/N, 128C/N, 129C/N, 130C/N, 131) ²⁸. Those mass differences in the reporter can be detected using high resolution mass spectrometry. Benefits of the TMT 10-plex comprise increased sample multiplexing for relative quantitation, better sample throughput and fewer missing quantitative channels among samples ²⁹. However, the use of TMT-based quantitative proteomics for Anisakidae investigations has not yet been applied.

Therefore, in this work we presented for the first time the proteome profiling of L3 and L4 *A. simplex* larvae by TMT-based quantitative proteomics. The methodology is based on the following steps: (a) protein extraction of L3 and L4 larvae, (b) HIFU-assisted trypsin digestion, (c) TMT-isobaric mass tag labeling following high-pH reversed-phase fractionation, and (d) LC-MS/MS analysis in a LTQ-Orbitrap Elite mass spectrometer.

The results from this study will provide a significant novel protein repository of L3 and L4 *A. simplex* larvae that will be very useful for the comparison of L3 and L4 development stages and for further anisakiosis investigations.

2. MATERIALS AND METHODS

2.1 *Anisakis simplex*

Parasitic nematodes were collected from the Biobank platform implemented for the PARASITE project (www.parasite-project.eu) at IIM-CSIC, Vigo, Spain. The experiments were performed on the L3 and L4 larval stages of *A. simplex* which were isolated: L3 from hake (*Merluccius merluccius*) and L4 from striped dolphin (*Stenella coeruleoalba*). All used larvae were taxonomically identified using conventional PCR to amplify the ITS region and Cox2 gene³⁰.

2.2 Protein extracts

Protein extracts were obtained from ten L3 and three L4 *A. simplex* larvae. For that, parasites were crushed manually with a sterile plastic pestle in 2 mL centrifuge tubes. Then, protein extraction was performed by homogenizing 0.2 g of sample in 1.5 mL of lysis buffer (60 mM Tris-HCl pH 7.5, 1% lauryl-maltoside, 5 mM PMFS and 1% DTT) on ice for 3 cycles of 30 s pulses in a sonicator device (IKA-Werke, Staufen, Germany). Proteins extracts were centrifuged at 16,000 g for 30 min at 4°C in a J221-M centrifuge (Beckman, CA, USA). The supernatant proteins were quantified using the bicinchoninic acid method (Pierce BCA Protein Assay Kit, Thermo Fisher Scientific, San Jose, CA, USA). All analyses were performed in triplicate.

2.3 SDS-Polyacrylamide Gel Electrophoresis

Protein extracts were separated and evaluated on 12% (v/v) polyacrylamide gels (acrylamide/N, N'-ethylene-bis-acrylamide, 200:1) with a stacking gel of 4% polyacrylamide. A total of 25 µg of proteins in Laemmli buffer were boiled for 5 min at

100 °C and separated per well in a Mini-PROTEAN 3 cell (Bio-Rad, Hercules, CA, USA). The running buffer consisted of an aqueous solution, composed by 1.44% (w/v) glycine, 0.67% Tris-base, and 0.1% SDS. Running conditions were 80 V for the first 20 min and then 120 V until the end of the electrophoresis. Gels were silver stained using the Pierce Silver Stain for Mass Spectrometry kit (Thermo Fisher Scientific) (Supplemental Figure 1).

2.4 Tryptic digestion by HIFU

Protein supernatants (100 µg) were ultrafast in-solution digested with 2.5 µg of trypsin (Promega, WI, USA) in 25 mM ammonium bicarbonate pH 8, with the simultaneous application of HIFU ³¹. The trypsin digestion was carried out as described before by Carrera et al. ³².

2.5 TMT peptide labeling

TMT 10-plex isobaric label reagents (0.8 mg Thermo Fisher Scientific) were re-suspended in 41 µL of anhydrous acetonitrile and added to 100 µg of protein digest dissolved in 100 µL of 0.1 M triethylammonium bicarbonate (TEAB) buffer solution. A total of two different complete experiments of Tandem Mass Tag (TMT) 10plex labelings (Thermo Fisher Scientific) with different biological replicates were performed. Within each experiment, samples were labeled with TMT10-plex in quadruplicate (L3 larvae: 126, 127N, 127C, 128N; L4 larvae: 128C, 129N, 129C, 130N). Additionally, a standard sample resulting of mixing equal amount of proteins for the eight samples was included (channel 131). Reaction was carried out by 1 hour at room temperature. To quench the reaction 8 µL of 5% hydroxylamine was added to each sample and incubated for 15 min. Samples were combined in a new tube at equal amounts. TMT-labeled

peptide concentration was measured using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific).

2.6 High-pH reversed-phase peptide fractionation

To increase the number of peptide identifications, labeled samples were fractionated using Pierce High-pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific) following manufacturer's instructions. After conditioning of the spin columns, a total of 100 µg of the total labelled and digested sample was dissolved in 300 µL of 0.1% TFA solution. Sample solution was loaded onto the spin column and centrifuged at 3000 x g for 2 min. Eluate was retained as "flow-through" fraction. The same was performed with Milli-Q water to retain the "wash" fraction. TMT-labeled samples require an additional column wash with 300 µL of 5% ACN, 0.1% trimethylamine (TEA) to remove unreacted TMT reagent. Fractionation of six different sample fractions were performed by stepwise using the appropriate elution solutions according to manufacturer's instructions. Peptide content of each fraction was determined by colorimetric analysis using the Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific) and evaporated to dryness using vacuum centrifugation (SpeedVac concentrator, Thermo Fisher Scientific). The samples (6 fractions plus the wash and flow-throughput) were stored at -80°C for further analysis.

2.7 LC-MS/MS analysis

Peptide samples were acidified with 0.1% formic acid, cleaned on a C₁₈ MicroSpin™ column (The Nest Group, South-borough, MA, USA) and analyzed by LC-MS/MS using a Proxeon EASY-nLC II liquid chromatography system (Thermo Fisher Scientific) coupled to a LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher

Scientific). Peptide separation (1 μ g) was done on a RP column (EASY-Spray column, 50 cm $\times$ 75 μ m ID, PepMap C18, 2 μ m particles, 100 Å pore size, Thermo Fisher Scientific) with a 10mm pre-column (Accucore XL C18, Thermo Fisher Scientific) using 0.1% formic acid and 98% ACN with 0.1% formic acid as mobile phases A and B, respectively. The 240 min of linear gradient from 5 to 35% B, at a flow of 300 nL/min was used. A spray voltage of 1.95 kV and a capillary temperature of 275°C were used for ionization. The peptides were analyzed in positive mode from 400–1600 amu (1 μ scan), followed by 10 data-dependent HCD MS/MS scans (1 μ scans), using a normalized collision energy of 38% and an isolation width of 1.5 amu. Dynamic exclusion for 30s after the second fragmentation event was applied and unassigned charged ions were excluded from the analysis.

2.8 Processing of the mass spectrometry data

All the MS/MS spectra acquired were analyzed using SEQUEST-HT (Proteome Discoverer 2.1 package, Thermo Fisher Scientific) against a custom-made database containing *A. simplex* plus *Ascaris suum*, *Toxocara canis*, *Brugia malayi*, *Loa loa* and *Caenorhabditis elegans* UniProt/TrEMBL protein entries (released 2017_08; 528,050 entries). These organisms were selected according to phylogenetic similarity. The following restrictions were used: full tryptic cleavage with up to 2 missed cleavage sites and tolerances of 7 ppm Da for parent ions and 0.6 Da for MS/MS fragment ions. TMT-labeling (lysine and peptide n-terminus) and carbamidomethylation of cysteine were set as fixed modifications. The permissible variable modifications were methionine oxidation and acetylation of the N-terminus of the protein. The results were subjected to statistical analysis to determine the peptide false discovery rate (FDR) using a decoy

database and the Target Decoy PSM Validator algorithm³³. The FDR was kept below 1%.

2.9 Modulated proteins and statistical analysis

Relative quantification was performed using the Quantification Module and normalization against total peptide amount (Proteome Discoverer 2.1 package, Thermo Fisher Scientific). The comparison of the abundance of proteins in L3 or in L4 was performed according to Abundance Ratio: L3/L4 (Supplemental Data 1). Log2 of the Abundance ratio: L3/L4 was calculated. Negative values (≤ -1) mean that protein is more abundant in L4, and positive values (≥ 1) in L3. Those extreme ratios let us classified all the proteins as equally abundant in both stages (Log2 Abundance ratio: L3/L4 = $<-1;1>$), and those more abundant in one of the stage. After global quantification several filters were applied to the global quantification results to obtain the final list of the most modulated proteins: a) proteins quantified with at least two unique peptides, b) at least a 2 fold change in L3/L4 normalized ratios, c) proteins quantified in the two TMTs and d) after Kruskal-Wallis statistical test using the R commander software. Differences in the modulation were considered significant when p-value ≤ 0.01 (Supplemental Data 4).

2.10 Functional analysis

To analyze functions and the involvement in the common biological processes, the final list of non-redundant protein IDs obtained after global quantification was classified into three different categories of Gene Ontology (GO): biological processes, classes of proteins and molecular functions. Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to ascribe identified proteins to particular biological mechanisms and cellular pathways (the established criteria: p adjusted

<0.05). GO and KEGG enrichment analysis were performed using PANTHER Classification System (Protein ANalysis THrough Evolutionary Relationships, <http://pantherdb.org>) and DAVID 6.8 (Database for Annotation, Visualization and Integrated Discovery, <https://david.ncifcrf.gov/home.jsp>), respectively. The significantly enriched functional GO categories and KEGG pathways were reported by comparing the input data with the background of GO/KEGG annotations in the *Caenorhabditis elegans* genome, as the phylogenetically closest nematode available in both databases. GO categorization applied PANTHER overrepresentation test controlled with a Bonferroni correction test to correct the smallest *p*-value of GO terms.

2.11 Network analysis

Network analysis was performed submitting both protein datasets, L3 and L4 development stages of *A. simplex*, obtained after global quantification, to the STRING (Search Tool for the Retrieval of Interacting Genes) software (v10.5) (<https://string-db.org/>). Interactions have been identified through comparing the input data with the background of the *Caenorhabditis elegans* genome, as the phylogenetically closest nematode, available in the STRING software. Proteins were represented with nodes and the interactions with continuous lines to represent direct interactions (physical), while indirect ones (functional) were presented by interrupted lines. Cluster networks were created using the MCL inflation parameter, which is included in STRING website and a value of 3 was selected for all the analyses.

3. RESULTS AND DISCUSSION

3.1 Global proteome of *A. simplex* L3 and L4 development stages

In this study, the TMT-based quantitative proteomics experiments were used to create a reference proteome dataset for each of the two development stages of *A. simplex*, L3 (n= 80 samples) and L4 (n= 24) (Supplemental Data 1).

A total of 11412 non-redundant peptides corresponding to 2443 different proteins were identified (Table 1). According to the Abundance L3/L4 ratios during global quantification more abundant proteins for L3 (330 proteins) and for L4 (571 proteins) were identified and listed (Supplemental Data 1). Additionally, the results showed a high degree of overlap (equally abundant 1542 proteins) between L3 and L4 developmental stages (Table 1). To our knowledge, this is the major dataset of proteins of *A. simplex* for two development stages (L3 and L4) identified to date.

The final global datasets were subsequently investigated by gene ontology and network analysis to gather more functional insights.

3.2 Functional analysis

To understand molecular function and biological processes of the identified proteins that they are involved in, gene ontology (GO) term was performed by PANTHER classification system (Supplemental Data 2). Among biological processes (Figure 1a) the smallest differences are related to cellular and metabolic processes (GO:0009987, GO:0008152). The biggest differences between L3 and L4 are related to the biological adhesion (GO:0022610), multicellular organismal processes (GO:0032501), developmental processes (GO:0032502) and biological regulation

(GO:0065007). In all cases, except biological regulation, L3 has highest number of proteins related to that biological process (Figure 1a) (Supplemental Data 2).

Identified proteins were also divided according to their molecular function (Figure 3b). Analysis showed that in both stages, binding (GO:0005488) and catalytic activity (GO:0003824) are the most common molecular functions of the analyzed proteins. The most significant differences occurred between the L3 and L4 stages in terms of receptor activity (GO:0004872) and structural molecule activity (GO:005198). In L4 number of proteins which have receptor activity is higher than in L3, respectively 5.3% to 3.3% from all identified proteins. Different situation is with structural molecule activity proteins, where L3 has higher number of that type of proteins than L4, respectively 14.5% to 11.9% from all identified proteins (Figure 1b). Moreover, the identified proteins when divided into two *A. simplex* development stages, were assigned into specific protein classes according to the annotations of the proteins (Figure 4) (Supplemental Data 2).

According to PANTHER, 23 protein classes with smallest p -value ($p < 0,05$) were taken under consideration. In L4 stage was more proteins classified to ligases (PC00142), lyases (PC00144), oxidoreductases (PC00176) and transferases (PC00220) than in L3. However, in L3 more proteins were assigned to calcium-binding proteins (PC00060), signaling molecules (PC00207), cell adhesion molecules (PC00069) and cytoskeletal proteins (PC00085) (Figure 2) (Supplemental Data 2).

KEGG pathway analysis by comparing the input data with the background of the *C. elegans* genome by DAVID 6.8 (<https://david.ncifcrf.gov/home.jsp>) showed that most of identified proteins in both stages: L3 and L4 were involved in main metabolic pathways (cel01100), ribosome (cel03010), biosynthesis of antibiotics (cel01130),

carbon metabolism (cel01200) and oxidative phosphorylation (cel00190) (Supplemental Data 2).

3.3 Network analysis

To gain information on protein interaction network after global quantification of L3 and L4 development stages of *A. simplex*, possible interactions were analyzed using the STRING v10.0 software. Two separate network maps have been made for each *A. simplex* development stage (Figures 3 and 4). A total of 92 proteins for L3 and 157 proteins for L4 constituted very complex and strongly interactive networks, which were clustered to a specified MCL inflation parameter (MCL = 3). According to MCL clustering 21 nodes in L3 stage and 29 nodes in L4 stage were obtained, each of them depicted as a group of different color circles. Most of all analyzed proteins were mutual interactive or interactive with other proteins in those complex protein-protein interaction networks (Figures 3 and 4; Supplemental Data 3). Among them, 236 interactions in L3 and 594 interactions were shown in connection with co-expression, co-occurrence and due to the appearance of any information on the interactions between those proteins in different databases (Figures 3 and 4; Supplemental Data 3). The most complex nodes of those interaction networks in L3 were those related to energy metabolism (orange circles), regulation of muscle contraction (red circles), protein catabolic processes (dark purple circles), oxidative metabolism (light pink circles) and cuticle creation (dark blue circles) (Figure 3, Supplemental Data 3). In contrast, in the L4 stage most complex nodes were those related to citrate cycle (light pink circles), aminoacyl-tRNA biosynthesis (peach circles), mitochondrial membrane respiratory chain (red circles) and vesicle-mediated transport (turquoise circles) (Figure 4, Supplemental Data 3).

Global proteome quantification analysis of L3 and L4 development stages of *Anisakis simplex* showed that in those two stages most of the proteins are involved in very important pathways like those related to energy metabolism, oxidation, cellular transport, and signaling pathways.

3.4 Quantitative modulated proteins

A total of 2443 different proteins of *A. simplex* were identified. Proteins were divided between L3 and L4 stages, where 1542 proteins were classified for both of them. The global quantification results showed 330 and 571 more abundant proteins, identified with at least one unique peptide, for L3 and for L4 development stage, respectively. From them, 248 and 393 were quantified with at least two unique peptides. After sequentially applying the next three criteria (Figure 5), 8 proteins in L3 and 12 proteins in L4 were selected as proteomic signature of those specific development stages of *A. simplex*. The list of those proteins was shown in Table 2.

A detailed representation of the relative abundance of all quantified proteins is depicted in the scatter plot (Figure 6). When comparing normalized TMT abundances values from all the L3 and L4 samples for each protein it is shown that L3 and L4 larvae differs in protein composition and proportion. The proteins selected as the proteomic signature of each *A. simplex* development stage, were the most modulated proteins occurred in both development stages of the parasite, but in transition L3/L4, they were the most statistically significantly ($p\text{-value} \leq 0.01$) abundant in one of the stage (Figure 7; Supplemental Data 4).

Most of all modulated proteins belong to hydrolases with particular regard to peptidases (Table 2). A number of specific proteins for each development stage have been recognized as allergens and registered in the Allergome database³⁴. Among these

is known protein such a somatic paramyosin (Ani s 2). Endochitinase, a glycoside hydrolase from the chitinase class III group was also assigned as an allergen of *A. simplex*⁸ (Table 2).

Considering proteins that were homologous to known allergens, according to Allergome database, 3 of quantified modulated proteins could be potential novel allergens of *A. simplex*. The pseudocoelomic globin (AF009, Aller-gome database) assigned to L3 larvae, prolylcarboxy peptidase like protein 5 (AF061, Aller-gome database), and L-threonine 3-dehydrogenase, mit. (AF029, Aller-gome database), selected as one of the L4 specific proteins. First protein, pseudocoelomic globin, represents a group of globins found in nematode worms. Globins are heme-containing proteins involved in binding and/or transport of oxygen. The major groups of globins are hemoglobins and myoglobins from vertebrates, invertebrate globins, leghemoglobins from plants, and flavohemoproteins from bacteria, where invertebrates and vertebrates globins causes allergic reactions^{35, 36, 37}. Second protein belongs to prolyl oligopeptidase family. All allergens belonging to this family are dipeptidyl peptidases. They were identified in several species of the dermatophytic fungus *Trichophyton* genus³⁸ and in bee and wasp venoms³⁹. Third one protein, L-threonine 3-dehydrogenase, is a zinc-dependent dehydrogenase. It is closely related to zinc alcohol dehydrogenases, which are found in bacteria, mammals, plants, and in fungi⁴⁰, where one of the allergens from *Candida albicans* is an alcohol dehydrogenase⁴¹.

Analyzing only modulated proteins in STRING v10.0 software (Figure 8) after applying MCL clustering, expected number of edges was 4, where current set of proteins showed 9. This means that these proteins have more interactions among themselves than what would be expected for a random set of proteins of similar size, drawn from the genome. Such an enrichment indicates that the proteins are partially

biologically connected, as a group what is shown in the Figure 8. This means that function blocking or silencing of appropriate genes for these proteins, especially those with the highest number of protein-protein interactions (i.e. glutamate dehydrogenase; A0A0M3JI05 or neprylisin-2; A0A0M3JXL8) could be the basis for combating *Anisakis simplex* L3 and L4 larvae, due to disruption of metabolically important processes for both stages.

Proteins which are not connected with any or not shown have not been studied very much and that their interactions might not yet be known to STRING. On the other hand, interactions have been identified through comparing the input data with the background of the *C. elegans* genome, the phylogenetically closest nematode to *A. simplex* available in the STRING software. However, it is free-living species, not parasitic one.

Obtained stage-specific proteins, could be used as targets to control/eliminate this parasite and in future eradicate the anisakiosis disease. Further detailed *in vitro* studies on the silencing of expression or blocking of activity, as well as post-translational modifications of these proteins, due to their high participation in important metabolic processes in Nematoda, are needed. Their accurate knowledge and description of the mechanism of action could be used in research on antiparasitic substances and give an answer to the question how to neutralize invasive L3 larvae and thus help in the fight with anisakiosis.

In a next step, it will be advantageous to verify our findings by using fractionated or recombinant proteins for the immunoblotting experiments, and to study the allergenicity of selected allergen candidates in more detail. On the other hand, determination of marker peptides of selected modulated *A. simplex* proteins may be used for the classification of food and feed contaminants.

4. CONCLUSIONS

In this study, we performed comparative quantitative proteomics analyses employing L3 and L4 *Anisakis simplex* development stages to reveal global proteomes of those two. The proteomes were analyzed using TMT label-based multiplex quantification analysis technique, which was based on the following steps: (a) protein extraction, (b) HIFU-assisted trypsin digestion, (c) TMT-isobaric mass tag labeling, and (d) global proteomic comparison analysis between two developmental stages of *A. simplex* using a LTQ-Orbitrap Elite mass spectrometer. Interestingly, 2443 different proteins were identified, where the results showed a high degree of overlap (1542 different proteins) among L3 and L4 of *A. simplex* and high amount of proteins more abundant for L3 (330) or L4 (571). The final list of modulated proteins consists 8 proteins in L3 and 12 proteins in L4 selected as proteomic signature of those specific development stages of *A. simplex*.

To our knowledge, the current study represents the most extensive analysis of global proteomes for *A. simplex* two development stages: L3 and L4, provided to date and might be very useful for further investigations under anisakiosis, including the determination of marker peptides of selected modulated *A. simplex* proteins what may be used for the classification of food and feed contaminants.

REFERENCES

- [1] N. E. Nieuwenhuizen. Anisakis – immunology of a foodborne parasitosis. *Parasite Immunol.* 38 (2016) 548-557.
- [2] C. JongYil, K.D. Murrell, A.J. Lymbery. Fish-borne parasitic zoonoses: status and issues. *Int. J. Parasitol.* 35 (2005) 1233–1254.
- [3] M.J. Rosales, C. Mascaró, C. Fernandez, F. Luque, M.S. Moreno, L. Parras, et al. Acute intestinal anisakiasis in Spain: a fourth-stage *Anisakis simplex* larva. *Mem Inst. Oswaldo Cruz.* 94 (1999) 823–826.
- [4] M. Uña-Gorospe, I. Herrera-Mozo, M.L. Canals, G. Martí-Amengual, P. Sanz-Gallen. Occupational disease due to *Anisakis simplex* in fish handlers. *Int. Marit Health.* 69 (2018) 264-269.
- [5] S. Klimpel, H.W. Palm, Anisakid nematode (Ascaridoidea) life cycles and distribution: increasing zoonotic potential in the time of climate change? In: Mehlhorn H. (eds) *Progress in Parasitology. Parasitology Research Monographs*, Springer, Berlin, Heidelberg. 2 (2011) 201-222.
- [6] J. Bratney, K.J. Clark. Effect of temperature on egg hatching and survival of larvae of *Anisakis simplex* B (Nematoda: Ascaridoidea). *Can. J. Zool.* 70 (1992) 274-279.
- [7] F.J. Baird, R.B. Gasser, A. Jabbar, A.L. Lopata. Foodborne anisakiasis and allergy. *Mol. Cell. Probes* 28 (2014) 167–174.
- [8] Ch.K. Fæste, K.R. Jonscher, M.M.W.B. Dooper, W. Egge-Jacobsen, A. Moen, A. Daschner, E. Egaas, U. Christians. Characterization of potential novel allergens in the fish parasite *Anisakis simplex*. *Proteomics.* 4 (2014) 140-155.
- [9] M.T. Audicana, M.W. Kennedy. *Anisakis simplex*: from obscure infectious worm to inducer of immune hypersensitivity. *Clin. Microbiol. Rev.* 21 (2008) 360-379.
- [10] J. Ivanović, M. Ž. Baltić, M. Bošković, N. Kilibarda, M. Dokmanović, R. Marković, J. Janjić, B. Baltić. Anisakis allergy in human. *Trends Food Sci. Tech.* 59 (2017) 25-29.
- [11] A.H. Lin, E. Florvaag, T. van Do, S.G.O. Johansson, A. Levsen, K. Vaali. IgE sensitization to the fish parasite *Anisakis simplex* in a norwegian population: A pilot study. *Scand. J. Immunol.* 75 (2015) 431-435.

- [12] S. Mattiucci, P. Cipriani, S. C. Webb, M. Paoletti, F. Marcer, B. Bellisario, D. I. Gibson, G. Nascetti. Genetic and morphological approaches distinguish the three sibling species of the *Anisakis simplex* species complex, with a species designation as *Anisakis berlandi* n. sp. for *A. simplex* sp. C (Nematoda: Anisakidae). *J. Parasitol.* 100 (2014) 199-214.
- [13] M.-F. Jeebhay, T.-G. Robins, S.-B. Lehrer, A. L. Lopata. Occupational seafood allergy: a review. *Occup. Environ. Med.* 58 (2001) 553-562.
- [14] M.T. Audicana, I.J. Ansotegui, L.F. Corres, M.W. Kennedy. *Anisakis simplex*: dangerous—dead and alive? *Trends Parasitol.* 18 (2002) 20-25.
- [15] A. M. Anadón, F. Romarís, M. Escalante, E. Rodríguez, T. Gárate, C. Cuéllar, F. M. Ubeira. The *Anisakis simplex* Ani s 7 major allergen as an indicator of true *Anisakis* infections. *Clin. Exp. Immunol.* 156 (2009) 471-478.
- [16] EFSA Panel on Biological Hazards (BIOHAZ). Scientific opinion on risk assessment of parasites in fishery products. *EFSA J.* 8 (2010) 1543.
- [17] M. Bao, G.J. Pierce, S. Pascual, M. González-Muñoz, S. Mattiucci, I. Mladineo, P. Cipriani, I. Bušelić, N.J.C. Strachan. Assessing the risk of an emerging zoonosis of worldwide concern: Anisakiasis. *Sci. Rep.* 7 (2017) 43699.
- [18] T. Baptista-Fernandes, M. Rodrigues, I. Castro, P. Paixão, P. Pinto-Marques, L. Roque, S. Belo, P. Manuel Ferreira, K. Mansinho, C. Toscano. Human gastric hyperinfection by *Anisakis simplex*: A severe and unusual presentation and a brief review. *Int. Infect. Dis.* 64 (2017) 38-41.
- [19] M. Llarena-Reino, E. Abollo, M. Regueira, H. Rodríguez, S. Pascual. Horizon scanning for management of emerging parasitic infections in fishery products. *Food Control.* 49 (2015) 49-58.
- [20] Commission Regulation (EU) No 1276/2011 of 8 December 2011 amending Annex III to Regulation (EC) No 853/2004 of the European Parliament and of the Council as regards the treatment to kill viable parasites in fishery products for human consumption, *Official Journal of the European Union* 327, 9.12.2011, p. 39–41.
- [21] M. Carrera, J. M. Gallardo, S. Pascual, Á. F. González, I. Medina. Protein biomarker discovery and fast monitoring for the identification and detection of Anisakids by parallel reaction monitoring (PRM) mass spectrometry. *J. Proteom.* 142 (2016) 130-137.
- [22] A. Cifuentes. Food analysis and foodomics. *J. Chromatogr. A.* 1216 (2009) 7109.

- [23] M. Herrero, C. Simó, V. García-Cañas, E. Ibáñez, A. Cifuentes. Foodomics: MS-based strategies in modern food science and nutrition. *Mass Spectr. Rev.* 31 (2012) 49-69.
- [24] J.M. Gallardo, I. Ortea, M. Carrera. Proteomics and its applications for food authentication and food-technology research. *Trend Anal Chem.* 52 (2013) 135-141.
- [25] M. Carrera, B. Cañas, J.M. Gallardo. Rapid direct detection of the major fish allergen, parvalbumin, by selected MS/MS Ion Monitoring mass spectrometry. *J. Proteom.* 75 (2012) 3211-3220.
- [26] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A.K. Mohammed, C. Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75 (2003) 1895-1904.
- [27] G.C. McAlister, E. L. Huttlin, W. Haas, L. Ting, M.P. Jedrychowski, Rogers, J. C.; K. Kuhn, I. Pike, R.A. Grothe, J.D. Blethrow, S.P. Gygi. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* 84 (2012) 7469-7478.
- [28] N. Rauniyar, J. R. Yates. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J. Proteome Res.* 13 (2014) 5293-5309.
- [29] T. Werner, I. Becher, G. Sweetman, C. Doce, M.M. Savitski, M. Bantscheff. High-resolution enabled TMT 8-plexing. *Anal. Chem.* 84 (2012) 7188-7194.
- [30] X. Zhu, S. D'Amelio, L. Paggi, R.B. Gasser. Assessing sequence variation in the internal transcribed spacers of ribosomal DNA within and among members of the *Contracaecum osculatum* complex (Nematoda: Ascaridoidea: Anisakidae). *Parasitol. Res.* 86 (2000) 677-683.
- [31] D. López-Ferrer, J.L. Capelo, J. Vázquez. Ultra-Fast Trypsin Digestion of Proteins by High Intensity Focused Ultrasound. *J. Proteom Res.* 4 (2005) 1569-1574.
- [32] M. Carrera, B. Cañas, D. López-Ferrer, C. Piñeiro, J.M. Gallardo. Fast monitoring of species-specific peptide biomarkers using high intensity focused ultrasound assisted tryptic digestion and selected MS/MS ion monitoring. *Anal. Chem.* 83 (2011) 5688-5695.
- [33] L. Kall, J. Canterbury, J. Weston, W.S. Noble, M.J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods.* 4 (2007) 923-925.

- [34] A. Mari, C. Rasi, P. Palazzo, E. Scala. Allergen databases: current status and perspectives. *Curr Allergy Asthma Rep.* 9 (2009) 376-83.
- [35] V. van Kampen, V. Liebers, A. Czuppon, X. Baur. Chironomidae hemoglobin allergy in Japanese, Swedish, and German populations. *Allergy.* 49 (1994) 9-12.
- [36] M.M. Fuentes, R. Palacios, M. Garces, M.L. Caballero, I. Moneo. Isolation and characterization of a heat-resistant beef allergen: myoglobin. *Allergy.* 59 (2004) 327-31.
- [37] J.T. Lecomte, D.A. Vuletich, A.M. Lesk. Structural divergence and distant relationships in proteins: evolution of the globins. *Curr Opin Struct Biol.* 15 (2005) 290-301.
- [38] J.A. Woodfolk, L.M. Wheatley, R.V. Piyasena, D.C. Benjamin, T.A. Platts-Mills. Trichophyton antigens associated with IgE antibodies and delayed type hypersensitivity. Sequence homology to two families of serine proteinases. *J Biol Chem.* 273 (1998) 29489-96.
- [39] S. Blank, H. Seismann, B. Bockisch, I. Braren, L. Cifuentes, M. McIntyre, D. Ruhl, J. Ring, R. Bredehorst, M.W. Ollert, T. Grunwald, E. Spillner. Identification, recombinant expression, and characterization of the 100 kDa high molecular weight Hymenoptera venom allergens Api m 5 and Ves v 3. *J Immunol.* 184 (2010) 5403-13.
- [40] H.W. Sun, B.V. Plapp. Progressive sequence alignment and molecular evolution of the Zn-containing alcohol dehydrogenase family. *J Mol Evol.* 34 (1992) 522-35.
- [41] H.D. Shen, K.B. Choo, H.H. Lee, J.C. Hsieh, W.L. Lin, W.R. Lee, S.H. Han. The 40-kilodalton allergen of *Candida albicans* is an alcohol dehydrogenase: molecular cloning and immunological analysis using monoclonal antibodies. *Clin Exp Allergy.* 21 (1991) 675-81.

FIGURE CAPTIONS

Figure 1. Biological processes (A) and molecular functions (B) of L3 and L4 *Anisakis simplex* development stages classified by PANTHER. Analysis performed on proteins obtained after global quantification.

Figure 2. Classes of proteins categorized per each L3 and L4 *Anisakis simplex* development stage by PANTHER. Analysis performed on proteins obtained after global quantification.

Figure 3. Protein-protein interaction network analysis of L3 *Anisakis simplex* development stage using STRING. Analysis performed on proteins obtained after global quantification.

Figure 4. Protein-protein interaction network analysis of L4 *Anisakis simplex* development stage using STRING. Analysis performed on proteins obtained after global quantification.

Figure 5. Summary of the filtering process during the selection of the most modulated proteins in L3 and L4 *Anisakis simplex* larvae.

Figure 6. Scatter plot with detailed representation of the relative abundance of all quantified proteins. Dots representing the modulated proteins in L3 and L4 *Anisakis simplex* marked with red with protein number. For the legend see Table 2.

Figure 7. The modulated proteins selected as the proteomic signature of each *A. simplex* development stage. Each protein L3/L4 transition is showed. Box-plots

representing ranks calculated for normalized TMT values for all L3 and L4 samples transformed into log2 values. Differences in the modulation were considered significant when $p\text{-value} \leq 0.01$.

Figure 8. Protein-protein interaction network analysis of *Anisakis simplex* the most modulated proteins using STRING. Edge confidence explained in the figure. For the legend see Table 2.

SUPPORTING INFORMATION

Supplemental Figure 1: The SDS-Polyacrylamide Gel electrophoresis. Protein extracts were separated and evaluated on 12% (v/v) polyacrylamide gels (200:1) with a stacking gel of 4% polyacrylamide. Gel was silver stained. 1-molecular weight marker; 2-5 L3 extracts; 6-9 L4 extracts.

Supplemental Data 1: Data repositories for the proteomes of L3 and L4 *Anisakis simplex* developmental stages.

Supplemental Data 2: Data repositories for the functional analysis of L3 and L4 *Anisakis simplex* development stages

Supplemental Data 3: Data repositories for the network analysis of L3 and L4 *Anisakis simplex* development stages using the String v10.0 software.

Supplemental Data 4: Data repository for the statistical analysis of L3 and L4 *Anisakis simplex* modulated proteins.

ACKNOWLEDGEMENTS

We are grateful to Lorena Barros Tajés (IIM-CSIC, Vigo, Spain) for her excellent practical assistance during the experiments. Thanks to PARASITE Biobank (www.parasite-project.eu) at Marine Research Institute (IIM-CSIC, Vigo, Spain) for the provision of the *Anisakis simplex* larvae used in this study. This work was supported by the Ramón Areces Foundation (XVII National grant), and as part of the project "Development Program of the University of Warmia and Mazury in Olsztyn" no. POWR.03.05.00-00-Z310 / 17.

Table 1. Number of identified and characterized proteins for *Anisakis simplex* using TMT10-plex labeling.

	Identified in this study	Characterized to date among Identified *	New characterized in this study	Still uncharacterized among Identified *
L3	330	40	241	49
L4	571	51	356	164
common	1542	202	1075	265
Total	2443	293	1672	478
SUM: 2443				

*According to UniProt/TrEMBL database for *Anisakis simplex* released 2017_08; 528.050 entries.

Table 2. The list of quantitative modulated proteins in *Anisakis simplex* specific for L3 and L4 larvae.

<i>Development stage</i>	<i>Protein name</i>	<i>UniProtKB/TrEMBL ID for Anisakis simplex</i>	<i>Gene</i>
L3	Pseudocoelomic globin	A0A0M3JEL6	<i>glb-1</i>
	Phosphoenolpyruvate carboxy kinase GTP	A0A0M3KCB9	<i>PEPCK</i>
	Endochitinase 1	A0A0M3K6E2	<i>CHIA1</i>
	Paramyosin	A0A158PP35	<i>unc-15</i>
	Putative leucine-rich repeat containing protein	A0A0M3J2S6	N/A
	Uncharacterized protein	A0A0M3K064	N/A
	Uncharacterized protein	A0A0M3J3L5	N/A
	Uncharacterized protein	A0A0M3K1G1	N/A
L4	Glutamate dehydrogenase	A0A0M3JI05	<i>gdh-1</i>
	Neprilysin-2	A0A0M3JXL8	<i>nep-2</i>
	Aminopeptidase N	A0A0M3K8U2	<i>ANPEP</i>
	C-type lectin protein	A0A0M3KDB5	<i>clec-64</i>
	Prolylcarboxy peptidase like protein 5	A0A0M3K3Q4	<i>pcp-5</i>
	L-threonine 3-dehydrogenase, mit.	A0A0M3K9T7	<i>tdh</i>
	Phosphotransferase	A0A0M3J2N1	N/A
	Glyoxylate reductase/ hydroxy pyruvate reductase	A0A0M3KK68	<i>grhpr</i>
	Uncharacterized protein	A0A0M3JCR4	N/A

Uncharacterized protein	A0A0M3K6J1	N/A
Uncharacterized protein	A0A0M3K2A5	N/A
Uncharacterized protein	A0A0M3IYI5	N/A

Note. N/A- not available, mit.- mitochondrial

Figure 1

Figure 2

Figure 4

Figure 5

L3: 330 proteins;

L4: 571 proteins

Quantificated with at least
two unique peptides

L3: 248 proteins;

L4: 393 proteins

Minimum 2 fold change
in L3/L4 normalized
ratios

L3: 74 proteins;

L4: 100 proteins

Quantificated in the two TMTs

L3: 34 proteins

L4: 69 proteins

p-value ≤ 0.01 after
Kruskal-Wallis statistical
test

Proteomic Signature of *Anisakis simplex*

L3: 8 PROTEINS;

L4: 12 PROTEINS

Figure 6.

Figure 7.

Figure 8.

Graphical Abstract

Table 1. Number of identified and characterized proteins for *Anisakis simplex* using TMT10-plex labeling.

	Identified in this study	Characterized to date among Identified *	New characterized in this study	Still uncharacterized among Identified *
L3	330	40	241	49
L4	571	51	356	164
common	1542	202	1075	265
Total	2443	293	1672	478
SUM: 2443				

*According to UniProt/TrEMBL database ANISI for *Anisakis simplex* released 2017_08; 528.050 entries.

Table 2. The list of quantitative modulated proteins in *Anisakis simplex* L3 and L4 larvae.

Development stage	Protein name	UniProtKB/TrEMBL ID for <i>Anisakis simplex</i>	Gene
L3	Pseudocoelomic globin	A0A0M3JEL6	<i>glb-1</i>
	Phosphoenolpyruvate carboxykinase GTP	A0A0M3KCB9	<i>PEPCK</i>
	Endochitinase 1	A0A0M3K6E2	<i>CHIA1</i>
	Paramyosin	A0A158PP35	<i>unc-15</i>
	Putative leucine-rich repeat containing protein	A0A0M3J2S6	N/A
	Uncharacterized protein	A0A0M3K064	N/A
	Uncharacterized protein	A0A0M3J3L5	N/A
	Uncharacterized protein	A0A0M3K1G1	N/A
L4	Glutamate dehydrogenase	A0A0M3JI05	<i>gdh-1</i>
	Neprilysin-2	A0A0M3JXL8	<i>nep-2</i>
	Aminopeptidase N	A0A0M3K8U2	<i>ANPEP</i>
	C-type lectin protein	A0A0M3KDB5	<i>clec-64</i>
	Prolylcarboxy peptidase like protein 5	A0A0M3K3Q4	<i>pcp-5</i>
	L-threonine 3-dehydrogenase, mit.	A0A0M3K9T7	<i>tdh</i>
	Phosphotransferase	A0A0M3J2N1	N/A
	Glyoxylate reductase/ hydroxypyruvate reductase	A0A0M3KK68	<i>grhpr</i>
	Uncharacterized protein	A0A0M3JCR4	N/A
	Uncharacterized protein	A0A0M3K6J1	N/A
	Uncharacterized protein	A0A0M3K2A5	N/A
	Uncharacterized protein	A0A0M3IYI5	N/A

Note. N/A- not available, mit.- mitochondrial

SIGNIFICANCE

A. simplex is a fish-borne parasite responsible for the human anisakiosis and allergic reactions around the world. The work describes for the first-time the comparison of the proteome of two *A. simplex* stages (L3 and L4). The strategy is based on four steps: (i) protein extraction, (ii) ultra-fast trypsin digestion under High-Intensity Focused Ultrasound (HIFU), (iii) TMT-isobaric mass tag labeling followed by high-pH reversed-phase fractionation and (iv) peptide analysis using a LTQ-Orbitrap Elite mass spectrometer. The workflow allows to select the most modulated proteins as proteomic signature of those specific development stages (L3 and L4) of *A. simplex*. Obtained stage-specific proteins, could be used as targets to control/eliminate this parasite and in future eradicate the anisakiosis disease.

Highlights

1. Proteomic analysis of two *Anisakis simplex* stages (L3 and L4) for the first time
2. Obtained stage-specific proteins, as a target to control/eliminate this parasite and in future eradicate the disease
3. Demonstrated (TMT)-based quantitative proteomics as a universal tool to analyze global proteomic dynamics in parasitic nematode stages